

Bayesian over-dispersed Poisson model and the Bornhuetter & Ferguson claims reserving method

Peter D. England

Towers Watson, London, UK

Richard J. Verrall

Cass Business School, City University, London, UK

Mario V. Wüthrich*

ETH Zurich, RiskLab, Department of Mathematics, Switzerland

Abstract

We consider the Bayesian over-dispersed Poisson (ODP) model for claims reserving in general insurance. We choose two different types of prior distributions for the parameters and then study the different Bayesian predictors. This study leads, on the one hand, to the classical chain ladder predictor and, on the other hand, to Bornhuetter & Ferguson predictors. We highlight (either analytically or numerically) how these predictors are obtained and how their prediction uncertainty can be determined.

Keywords

Claims reserving; Over-dispersed Poisson model; Bornhuetter and Ferguson method; Chain ladder method; Prediction uncertainty; Mean square error of prediction; Markov chain Monte Carlo method

1 Bayesian over-dispersed Poisson model

1.1 Introduction

A number of papers have appeared in the recent literature looking at stochastic models related to the Bornhuetter & Ferguson (BF, 1972) method of claims reserving, see for example Alai *et al.* (2009), Mack (2008), Saluz *et al.* (2011), Verrall (2004). The basic philosophy underlying these papers, and the BF method, is that there is external knowledge about the ultimate losses that is not contained in the runoff triangles of data. In statistical methodology, the usual way to incorporate such external knowledge is to use Bayesian methods. This paper examines the use of Bayesian methods for over-dispersed Poisson (ODP) models. The Bayesian ODP model treated in this paper was briefly covered in England & Verrall (2002), the present paper provides a much more detailed analysis and examines the use of different prior distributions and posterior estimators. We provide analytical results, where possible, which allow for intuitive interpretations. Where it is not possible

*Correspondence to: Mario V. Wüthrich, ETH Zurich, RiskLab, Department of Mathematics, 8092 Zurich, Switzerland. E-mail: mario.wuethrich@math.ethz.ch

to derive analytical results, we use Markov chain Monte Carlo (MCMC) methods to obtain numerical results.

Although this paper is related to the BF method, in that the underlying philosophy is similar, the results are not the same as the results from applying the conventional BF method. The reason for this is that the BF method uses the same runoff pattern as the chain ladder (CL) technique, whereas the application of Bayesian prior distributions to the rows of the claims development triangle naturally affects the posterior distributions of the parameters for the columns (i.e. the runoff parameters) of the claims development triangle – even if non-informative prior distributions are used for the latter. It is possible to construct a Bayesian BF model where the runoff pattern is exactly the same as in the CL technique, see Verrall (2004), but in that case it is not clear that this is a statistically optimal estimator. Thus, the purpose of this paper is to examine Bayesian models which incorporate prior knowledge about ultimate losses (as the BF method does), but it is *not* the purpose to reproduce the results from the BF method exactly, as in Alai *et al.* (2009).

The model assumptions are set out in full in Sections 1.2, 2.1 and 3.1, but the basic idea is to use an ODP model for the incremental claims with cross-classified means $\mu_i\gamma_j$, where μ_i is the row parameter in accident year i (related to the exposure of accident year i) and γ_j is the column parameter for development period j (related to the runoff pattern), and to apply prior distributions to these parameters. We will assume that there is no prior knowledge about the runoff parameters, and we use non-informative prior distributions for γ_j . By assuming informative prior distributions for the μ_i 's we can incorporate external knowledge about the ultimate losses. We investigate a number of different formulations of these informative prior distributions, and examine the properties of the resulting posterior estimators. We also compare our results with the traditional BF method.

An important observation will be that although we choose non-informative prior distributions for the parameters, their shapes may have a significant influence on the resulting claims reserves.

Organization of the paper. In the remainder of this section we define the general Bayesian ODP Model and we discuss prediction in a Bayesian framework. In Sections 2 and 3 we then specify two different types of prior distributions (the uniform prior model with log link function and the gamma prior model). Parameter estimates, e.g. for γ_j , are always denoted by $\hat{\gamma}_j$ in the uniform prior model with log link function and with $\hat{\gamma}_j^*$ in the gamma prior model. In Section 4 we discuss parameter estimation via simulation methods and in Section 5 a numerical example is provided. All the statements are proved in the appendix.

1.2 Model assumptions

The model assumptions are similar to those in the Bayesian claims reserving models presented in England & Verrall (2002, 2006), Verrall (2004) and Wüthrich & Merz (2008), Section 4.4. We assume that the parameters are modelled through prior distributions and, conditional on these parameters, the incremental claims $X_{i,j}$ have independent ODP distributions for accident years $i \in \{0, \dots, I\}$ and development years $j \in \{0, \dots, I\}$. The final development year is given by I and the observations at time I are given in the (upper) runoff triangle

$$\mathcal{D}_I = \{X_{i,j} : i + j \leq I\}.$$

Our goal is to predict the future claims in the lower triangle $\mathcal{D}_I^c = \{X_{i,j} : i + j > I, i \leq I\}$.

Model 1.1 (Bayesian ODP model)

- $\mu_0, \dots, \mu_I, \gamma_0, \dots, \gamma_I, \varphi$ are independent positive random variables with joint density $u(\cdot)$.
- Conditionally, given parameters $\Theta = (\mu_0, \dots, \mu_I, \gamma_0, \dots, \gamma_I, \varphi)$, X_{ij} are independent random variables with

$$\frac{X_{ij}}{\varphi} \Big|_{\Theta} \stackrel{(d)}{\sim} \text{Poi}(\mu_i \gamma_j / \varphi).$$

The parameter μ_i plays the role of the row parameter (related to the exposure of accident year i , see (1.2)), the γ_j 's describe column parameters (related to an incremental claims development (runoff) pattern that is not necessarily normalized, see (1.2)) and φ describes the dispersion parameter. We obtain the following first two conditional moments

$$E[X_{ij} | \Theta] = \mu_i \gamma_j \quad \text{and} \quad \text{Var}(X_{ij} | \Theta) = \varphi \mu_i \gamma_j, \quad (1.1)$$

and the conditional total ultimate claim of accident year i is given by

$$E\left[\sum_{j=0}^I X_{ij} \Big| \Theta\right] = \mu_i \sum_{j=0}^I \gamma_j. \quad (1.2)$$

We analyze the Bayesian ODP Model 1.1 for different types of prior distributions for Θ and different types of parameter estimates for Θ (see (1.3)–(1.4) below).

1.3 Bayesian predictors

Assume the Bayesian ODP Model 1.1 to hold. Using Bayes' Theorem we find the posterior density of Θ , given the observations \mathcal{D}_I , by

$$u(\theta | \mathcal{D}_I) \propto \prod_{i+j \leq I} \exp\left\{-\frac{\mu_i \gamma_j}{\varphi}\right\} \frac{\left(\frac{\mu_i \gamma_j}{\varphi}\right)^{X_{ij}}}{\left(\frac{X_{ij}}{\varphi}\right)!} u(\theta),$$

where the proportionality sign \propto means up to normalization w.r.t. the random vector Θ . In Bayesian theory there are two commonly used predictors, the minimum mean square error (MMSE) predictor and the maximum a posteriori (MAP) predictor for Θ , given \mathcal{D}_I . These are given by

$$\hat{\Theta}^{MMSE} = E[\Theta | \mathcal{D}_I], \quad (1.3)$$

$$\hat{\Theta}^{MAP} = \arg \max_{\theta} u(\theta | \mathcal{D}_I). \quad (1.4)$$

The predictor $\hat{\Theta}^{MMSE}$ minimizes the conditional prediction variance (see also (2.7) below) and the predictor $\hat{\Theta}^{MAP}$ is the maximum likelihood estimator (MLE) for the posterior density $u(\theta | \mathcal{D}_I)$. The MAP predictor $\hat{\Theta}^{MAP}$ has the advantage that it can often be calculated analytically. On the contrary, it has a bias term $\hat{\Theta}^{MMSE} - \hat{\Theta}^{MAP}$, relative to the posterior density $u(\theta | \mathcal{D}_I)$, that can, in general, only be calculated numerically, for example, using the MCMC methodology. This is discussed in the rest of this paper.

2 Uniform prior distributions and the chain ladder method

In this section we start with uniform priors and log links for the parameters μ_i and γ_j . Such a model has already been studied in England & Verrall (2006), Section 7.1. The crucial consequence of the

uniform priors assumption is that if we make them non-informative we obtain the classical CL estimate from the MAP predictors. In this spirit, this model is another example that replicates the CL reserves (see also Subsection 2.3).

2.1 The (non-informative) uniform priors model with log link

We define the parameters on the log scale: $\alpha_i = \log(\mu_i)$ and $\beta_j = \log(\gamma_j)$.

Model 2.1 In addition to Model 1.1 we assume that $\alpha_i = \log(\mu_i)$ are uniformly distributed on $(-m, m)$ for $m > 0$, and $\beta_j = \log(\gamma_j)$ are uniformly distributed on $(-b, b)$ for $b > 0$ and $\varphi > 0$ is constant.

Remark. It might be more appropriate to use different uniform priors for each parameter, e.g. α_i are uniformly distributed on $(-m_i, m_i)$. However, if we choose non-informative priors for $\alpha_i = \log(\mu_i)$ and $\beta_j = \log(\gamma_j)$ (i.e. we will let $m \rightarrow \infty$ and $b \rightarrow \infty$), then the specific prior differences between the α_i 's and between the β_j 's are not relevant.

With (1.1) we obtain

$$\log E[X_{i,j}|\Theta] = \log(\mu_i \gamma_j) = \alpha_i + \beta_j,$$

which illustrates the role of the log link function, see also England & Verrall (2002), Section 2.3. That is, with the log link function we derive the generalized linear model form.

The posterior density under Model 2.1 is given by

$$u(\theta|\mathcal{D}_I) \propto \prod_{i+j \leq I} \exp\left\{-\frac{e^{\alpha_i} e^{\beta_j}}{\varphi}\right\} (e^{\alpha_i} e^{\beta_j})^{\frac{x_{i,j}}{\varphi}} \prod_{i=0}^I \frac{1}{2m} 1_{(-m,m)}(\alpha_i) \prod_{j=0}^I \frac{1}{2b} 1_{(-b,b)}(\beta_j).$$

If we assume that m and b are sufficiently large (we comment on this below) then the MAP predictors for α_i and β_j can be found by maximizing the posterior log-likelihood function $\log u(\theta|\mathcal{D}_I)$ analytically, see Section 2.3. This provides MAP estimators $\hat{\alpha}_i$ and $\hat{\beta}_j$ for α_i and β_j , respectively, that correspond to the solution of the following system of equations, see also e.g. (2.16)–(2.17) in Wüthrich & Merz (2008),

$$e^{\alpha_i} \sum_{j=0}^{I-i} e^{\beta_j} = \sum_{j=0}^{I-i} X_{i,j}, \quad \text{for } i = 0, \dots, I, \quad (2.1)$$

$$e^{\beta_j} \sum_{i=0}^{I-j} e^{\alpha_i} = \sum_{i=0}^{I-j} X_{i,j}, \quad \text{for } j = 0, \dots, I. \quad (2.2)$$

Remarks 2.2

- $C_{i,j} = \sum_{k=0}^j X_{i,k}$ is called the cumulative claim of accident year i up to development year j . The (total) ultimate claim of accident year i is denoted by $C_{i,I}$ and the outstanding loss liabilities at time I for accident year i are given by

$$R_i = \sum_{j=I-i+1}^I X_{i,j} = C_{i,I} - C_{i,I-i}, \quad (2.3)$$

under the assumption that $X_{i,j}$ denotes claims payments. The final goal is to predict these outstanding loss liabilities R_i and to determine the prediction uncertainty.

- The solution to (2.1)–(2.2) is not unique, i.e. whenever $\hat{\alpha}_i$ and $\hat{\beta}_j$ solve the system (2.1)–(2.2), then for any $K \in \mathbb{R}$ also $\hat{\alpha}_i + K$ and $\hat{\beta}_j - K$ solve the system (2.1)–(2.2). The requirement m and b sufficiently large now means that there exists at least one $K \in \mathbb{R}$ such that the solution $(\hat{\alpha}_0 + K, \dots, \hat{\alpha}_I + K, \hat{\beta}_0 - K, \dots, \hat{\beta}_I - K)$ of (2.1)–(2.2) is within $[-m, m]^{I+1} \times [-b, b]^{I+1}$. We fix such a constant K and then denote the resulting MAP predictor by

$$\hat{\Theta}^{MAP} = \left(e^{\hat{\alpha}_0^{MAP}}, \dots, e^{\hat{\alpha}_I^{MAP}}, e^{\hat{\beta}_0^{MAP}}, \dots, e^{\hat{\beta}_I^{MAP}} \right).$$

The MAP predictor for the outstanding loss liabilities R_i in (2.3) is then defined by

$$\hat{R}_i^{MAP} = e^{\hat{\alpha}_i^{MAP}} \sum_{j=I-i+1}^I e^{\hat{\beta}_j^{MAP}}.$$

We see that K cancels in this product and hence the specific choice of K is not important as long as at least one such K exists.

- The MAP optimization problem (2.1)–(2.2) can be solved analytically. This is discussed in Section 2.3, below.
- For the priors of α_i we can either use informative priors (i.e. $m < \infty$) or non-informative priors (i.e. $m \rightarrow \infty$). However, since we have only one parameter, namely m , we always have prior expected value $E[\alpha_i] = 0$ and variance $\text{Var}(\alpha_i) = m^2/3$. Because we would like to have more flexibility in these parameter choices (if we have prior knowledge on α_i), we consider different priors in Section 3, which then leads to a Bayesian BF model. For the BF method we refer to Bornhuetter & Ferguson (1972).
- Note that the MAP predictors do not depend on the explicit choices of m , b and φ , once m and b are sufficiently large. On the other hand the MMSE predictors will depend on these parameter choices.

The MMSE predictor for R_i in (2.3) is given by

$$\begin{aligned} \hat{R}_i^{MMSE} &= E[R_i | \mathcal{D}_I] = E \left[\sum_{j=I-i+1}^I X_{ij} | \mathcal{D}_I \right] \\ &= \sum_{j=I-i+1}^I E \left[\mu_i \gamma_j | \mathcal{D}_I \right] = \sum_{j=I-i+1}^I E[e^{\alpha_i} e^{\beta_j} | \mathcal{D}_I]. \end{aligned} \quad (2.4)$$

Due to the posterior dependence between α_i and β_j , given \mathcal{D}_I , this cannot be further decoupled and calculated in closed form, see also Verrall (2004). Therefore, the MMSE predictor can only be calculated numerically.

We analyze the right-hand side of (2.4) in more detail. We denote $\alpha = (\alpha_0, \dots, \alpha_I)$, $\beta = (\beta_0, \dots, \beta_I)$, $\mu = (\mu_0, \dots, \mu_I)$, $\gamma = (\gamma_0, \dots, \gamma_I)$. Doing the following change of variables $\mu_i = e^{\alpha_i}$ and $\gamma_j = e^{\beta_j}$ we obtain

$$\begin{aligned} \hat{R}_{i,j}^{MMSE} &= E[e^{\alpha_i} e^{\beta_j} | \mathcal{D}_I] = \int_{\mathbb{R}^{2I}} e^{\alpha_i} e^{\beta_j} u(\alpha, \beta | \mathcal{D}_I) d\alpha d\beta \\ &= \int_{\mathbb{R}_+^{2I}} \mu_i \gamma_j \mu(\mu, \gamma | \mathcal{D}_I) \prod_{k=0}^I \frac{1}{\mu_k} \prod_{l=0}^I \frac{1}{\gamma_l} d\mu d\gamma, \end{aligned} \quad (2.5)$$

with posterior density

$$u(\mu, \gamma | \mathcal{D}_I) \prod_{k=0}^I \frac{1}{\mu_k} \prod_{l=0}^I \frac{1}{\gamma_l} \propto \prod_{i+j \leq I} \exp \left\{ -\frac{\mu_i \gamma_j}{\varphi} \right\} \left(\mu_i \gamma_j \right)^{\frac{x_{ij}}{\varphi}} \quad (2.6)$$

$$\times \prod_{i=0}^I \frac{\mu_i^{-1}}{2m} 1_{(e^{-m}, e^m)}(\mu_i) \prod_{j=0}^I \frac{\gamma_j^{-1}}{2b} 1_{(e^{-b}, e^b)}(\gamma_j).$$

Maximizing the right-hand side of (2.6) provides the MAP estimators $\hat{\mu}_i^{MAP}$ and $\hat{\gamma}_j^{MAP}$.

Remarks 2.3

- Basically the same remarks about the uniqueness of the MAP estimators $\hat{\mu}_i^{MAP}$ and $\hat{\gamma}_j^{MAP}$ apply as in Remarks 2.2: (i) they are only unique up to multiplication (and division, respectively) with a positive constant; (ii) we choose $m > 0$ and $b > 0$ so large that the mode of the density (2.6) lies within $[e^{-m}, e^m]^{I+1} \times [e^{-b}, e^b]^{I+1}$.
- The MAP optimization problem (2.6) can be solved analytically. This is discussed in Section 3.2, below.
- The MAP predictor for the outstanding loss liabilities R_i in (2.3) is then defined by

$$\check{R}_i^{MAP} = \hat{\mu}_i^{MAP} \sum_{j=I-i+1}^I \hat{\gamma}_j^{MAP}.$$

This now leads to a slightly unpleasant observation. Note that the MMSE predictor in (2.5) does not depend on the parametrization. This is not true for the MAP predictor! The MAP estimators $\hat{\alpha}_i^{MAP}$ and $\hat{\beta}_j^{MAP}$ solve the system of equations (2.1)–(2.2), whereas the MAP estimators $\hat{\mu}_i^{MAP}$ and $\hat{\gamma}_j^{MAP}$ will solve the system of equations (3.11)–(3.12), below. Because these two systems of equations differ, we find

$$e^{\hat{\alpha}_i^{MAP}} e^{\hat{\beta}_j^{MAP}} \neq \hat{\mu}_i^{MAP} \hat{\gamma}_j^{MAP} \quad \text{which in general implies} \quad \hat{R}_i^{MAP} \neq \check{R}_i^{MAP}.$$

This property is well known in Bayesian statistics, see for example Smith (1998). It gives us a first indication that the MAP predictor is not always suitable in a Bayesian context.

2.2 Prediction uncertainty

We measure the prediction uncertainty in terms of the *conditional mean square error of prediction* (MSEP) which for a \mathcal{D}_I -measurable predictor \hat{R}_i for R_i is given by, see also Section 3.1 in Wüthrich & Merz (2008),

$$\text{mse}_{R_i|\mathcal{D}_I}(\hat{R}_i) = E \left[(R_i - \hat{R}_i)^2 | \mathcal{D}_I \right] = \text{Var}(R_i | \mathcal{D}_I) + \left(\hat{R}_i^{MMSE} - \hat{R}_i \right)^2. \quad (2.7)$$

From (2.7) we see that the MMSE predictor $\hat{R}_i^{MMSE} = E[R_i | \mathcal{D}_I]$ minimizes the conditional MSEP. The conditional MSEP for the MAP predictor is given by

$$\begin{aligned} \text{mse}_{R_i|\mathcal{D}_I}(\hat{R}_i^{MAP}) &= E \left[(R_i - \hat{R}_i^{MAP})^2 | \mathcal{D}_I \right] \\ &= \text{mse}_{R_i|\mathcal{D}_I}(\hat{R}_i^{MMSE}) + \left(\hat{R}_i^{MMSE} - \hat{R}_i^{MAP} \right)^2 \geq \text{mse}_{R_i|\mathcal{D}_I}(\hat{R}_i^{MMSE}). \end{aligned} \quad (2.8)$$

The MMSE predictor and the conditional MSEP can, in general, only be determined numerically, using e.g. the MCMC methodology.

First conclusions. In many situations the MAP predictor \hat{R}_i^{MAP} has the advantage over the MMSE predictor \hat{R}_i^{MMSE} that it can be calculated analytically. The MMSE predictor on the other hand has the advantage that it minimizes the prediction uncertainty if we use the conditional MSEP as uncertainty measure. The MAP predictor obtains a positive bias term $\left(\hat{R}_i^{MMSE} - \hat{R}_i^{MAP}\right)^2$, see (2.8). This bias term however needs to be interpreted carefully: it is always measured w.r.t. the posterior density $u(\theta|\mathcal{D}_I)$.

2.3 Link to the chain ladder algorithm

The remarkable property of the MAP predictor \hat{R}_i^{MAP} in Model 2.1 with non-informative priors and log link is that it is equal to the CL reserves \hat{R}_i^{CL} . That is, the non-informative priors Bayesian Model 2.1 with MAP predictors is another stochastic model that leads to the CL reserves: since our system (2.1)–(2.2) of equations is exactly the same as the one for the ODP model, see (2.16)–(2.17) in Wüthrich & Merz (2008), we have

$$\hat{R}_i^{MAP} = \hat{R}_i^{CL}. \quad (2.9)$$

In the literature this was, for example, proved by Mack (1991). Therefore, we define for $j = 0, \dots, I-1$ the CL factor estimators

$$\hat{f}_j = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}}.$$

Corollary 2.18 and Remarks 2.19 in Wüthrich & Merz (2008) then imply that (for the appropriate normalizing constant K)

$$\hat{\beta}_j^{MAP} = \log \left[\left(1 - \frac{1}{\hat{f}_{j-1}}\right) \prod_{k=j}^{I-1} \frac{1}{\hat{f}_k} \right] \quad \text{and} \quad \hat{\alpha}_i^{MAP} = \log \left[\frac{\sum_{j=0}^{I-i} X_{i,j}}{\sum_{j=0}^{I-i} e^{\hat{\beta}_j^{MAP}}} \right]. \quad (2.10)$$

That is, we can explicitly calculate the MAP predictors. Moreover, this gives another stochastic model that allows for the calculation of the conditional MSEP given in (2.8). Unlike in Mack's (1993) distribution-free CL model and in the ODP model (see England & Verrall (2002), Section 7.2) we do not need any approximations here for the estimation of the MSEP, but we calculate the exact conditional MSEP value (2.8) numerically in this Bayesian inference model (using the MCMC methodology).

In this spirit the parameter uncertainty of the estimate Θ is part of the model, see (2.8). Moreover, because we have all key figures in terms of the full posterior distributions, we can calculate any risk measure of interest (not only the conditional MSEP).

3 Gamma prior distributions

In Section 2 we have used uniform priors with log links in order to obtain the CL reserves. In this section gamma prior distributions (with the identity link) are used, especially for the modelling of the row parameters μ_i . This allows us to incorporate prior expert knowledge about the model parameters and we obtain claims reserves in a similar spirit to the BF method. However, in our model, we still have the freedom to determine how much credibility weight we give to the prior knowledge. A similar

Bayesian ODP model with gamma priors has, for example, already been studied in Section 7.11 of England & Verrall (2002) and Example 4.51 of Wüthrich & Merz (2008).

3.1 Informative priors for the row parameters

Model 3.1 In addition to Model 1.1 we assume that μ_i are Γ -distributed with mean $m_i > 0$ and shape parameter $a_i > 0$, γ_j are Γ -distributed with mean $c_j > 0$ and shape parameter $b > 0$, and $\varphi > 0$ is constant.

In contrast to Model 2.1, we now extend the prior model for μ_i to a two-parameter distribution. Our aim is to keep the mean m_i fixed and study the sensitivity in the shape parameter a_i . The priors for γ_j will be chosen to be non-informative (i.e. $b \rightarrow 0$).

The posterior density (likelihood function) in Model 3.1 is given by

$$u(\theta|\mathcal{D}_I) \propto \prod_{i+j \leq I} \exp\left\{-\frac{\mu_i \gamma_j}{\varphi}\right\} \left(\mu_i \gamma_j\right)^{\frac{X_{ij}}{\varphi}} \prod_{i=0}^{I-1} \mu_i^{a_i-1} \exp\left\{-\frac{a_i \mu_i}{m_i}\right\} \prod_{j=0}^{I-1} \gamma_j^{b-1} \exp\left\{-\frac{b \gamma_j}{c_j}\right\}. \quad (3.1)$$

The MAP predictors using non-informative priors for γ_j (i.e. $b \rightarrow 0$) are then found by solving

$$\mu_i \left(\sum_{j=0}^{I-i} \gamma_j + \frac{a_i \varphi}{m_i} \right) = \sum_{j=0}^{I-i} X_{i,j}^* + a_i \varphi, \quad \text{for } i = 0, \dots, I, \quad (3.2)$$

$$\gamma_j \sum_{i=0}^{I-j} \mu_i = \sum_{i=0}^{I-j} X_{i,j}^*, \quad \text{for } j = 0, \dots, I, \quad (3.3)$$

with adjusted incremental claims

$$X_{i,j}^* = \begin{cases} X_{i,j} & \text{for } j \geq 1 \text{ and } i \geq 1, \\ X_{i,j} - \varphi & \text{for } (j = 0 \text{ and } i \geq 1) \text{ or } (j \geq 1 \text{ and } i = 0), \\ X_{i,j} + (I-1)\varphi & \text{for } j = 0 \text{ and } i = 0, \end{cases}$$

and adjusted cumulative claims $C_{i,j}^* = \sum_{k=0}^j X_{i,k}^*$.

Therefore, the MAP predictors for non-informative claims development pattern γ_j will be a function of the parameters φ , $\mathbf{m} = (m_0, \dots, m_I)$ and $\mathbf{a} = (a_0, \dots, a_I)$.

Lemma 3.2 We assume Model 3.1 is fulfilled, and we assume that $\sum_{j=0}^{I-i} X_{i,j}^* > 0$ for all $i = 0, \dots, I$ and $\sum_{i=0}^{I-j} X_{i,j}^* > 0$ for all $j = 0, \dots, I$. The solution to (3.2)–(3.3) satisfies $\mu_i > 0$ and $\gamma_j > 0$ for all $a_i \geq 0$ and $i, j = 0, \dots, I$.

We first state a CL type result. Note that in the following lemma we do a CL argument on the rows instead of on the columns; and for $\mathbf{a} = 0$ we obtain the CL method on rows. Its proof is similar to the classical CL result, see e.g. Section 2.4 in Wüthrich & Merz (2008).

Lemma 3.3 In Model 3.1 equations (3.2)–(3.3) imply for $j = 0, \dots, I-1$

$$\frac{\sum_{k=0}^{I-j} \mu_k}{\sum_{k=0}^{I-j-1} \mu_k} = \frac{\sum_{k=0}^{I-j} C_{k,j}^* + \varphi \sum_{k=0}^{I-j} a_k \left(1 - \frac{\mu_k}{m_k}\right)}{\sum_{k=0}^{I-j-1} C_{k,j}^* + \varphi \sum_{k=0}^{I-j-1} a_k \left(1 - \frac{\mu_k}{m_k}\right)}.$$

The statement of Lemma 3.3 can also be written in incremental form, i.e. for $i = 1, \dots, I$

$$\frac{\mu_i}{\sum_{k=0}^{i-1} \mu_k} = \frac{C_{i,I-i}^* + \varphi a_i \left(1 - \frac{\mu_i}{m_i}\right)}{\sum_{k=0}^{i-1} C_{k,I-i}^* + \varphi \sum_{k=0}^{i-1} a_k \left(1 - \frac{\mu_k}{m_k}\right)}. \quad (3.4)$$

This implies the following theorem:

Theorem 3.4 *In Model 3.1 equations (3.2)–(3.3) imply for $i = 1, \dots, I$*

$$\mu_i = \frac{C_{i,I-i}^* + a_i \varphi}{\sum_{k=0}^{i-1} C_{k,I-i}^* + \varphi \sum_{k=0}^{i-1} \mu_k \left(\frac{a_k}{\mu_k} - \frac{a_k}{m_k} + \frac{a_i}{m_i}\right)} \sum_{k=0}^{i-1} \mu_k, \quad (3.5)$$

and

$$\sum_{i=0}^I \frac{a_i \mu_i}{m_i} = \sum_{i=0}^I a_i. \quad (3.6)$$

Theorem 3.4 is now the key to obtain the MLE which are the same as the solutions of equations (3.2)–(3.3). Note that the right-hand side of (3.5) only depends on μ_0, \dots, μ_{i-1} . Therefore, once we know the initial value μ_0 , the remaining estimators for μ_1, \dots, μ_I are calculated iteratively by (3.5). This is discussed in more detail below.

Solution to (3.2)–(3.3) for $a \in (0, \infty)$. We apply Theorem 3.4. Choose an initial value $\tilde{\mu}_0(\mu) = \mu > 0$, then using (3.5) we define iteratively for $i = 1, \dots, I$

$$\tilde{\mu}_i(\mu) = \frac{C_{i,I-i}^* + a_i \varphi}{\sum_{k=0}^{i-1} C_{k,I-i}^* + \varphi \sum_{k=0}^{i-1} \tilde{\mu}_k(\mu) \left(\frac{a_k}{\tilde{\mu}_k(\mu)} - \frac{a_k}{m_k} + \frac{a_i}{m_i}\right)} \sum_{k=0}^{i-1} \tilde{\mu}_k(\mu) = \frac{1}{\mathcal{I}_i(\mu)} \sum_{k=0}^{i-1} \tilde{\mu}_k(\mu),$$

where we have defined

$$\mathcal{I}_i(\mu) = \left(\frac{C_{i,I-i}^* + a_i \varphi}{\sum_{k=0}^{i-1} C_{k,I-i}^* + \varphi \sum_{k=0}^{i-1} \tilde{\mu}_k(\mu) \left(\frac{a_k}{\tilde{\mu}_k(\mu)} - \frac{a_k}{m_k} + \frac{a_i}{m_i}\right)} \right)^{-1}.$$

Note that the vector $(\tilde{\mu}_0(\mu), \dots, \tilde{\mu}_I(\mu))$ is now a function of one single parameter $\mu > 0$. The MAP predictors for (3.2)–(3.3) are then found by using the normalizing condition (3.6), that is, choose $\mu > 0$ such that

$$\sum_{i=0}^I \frac{a_i \tilde{\mu}_i(\mu)}{m_i} = \sum_{i=0}^I \frac{a_i}{m_i} \frac{1}{\mathcal{I}_i(\mu)} \sum_{k=0}^{i-1} \tilde{\mu}_k(\mu) \stackrel{!}{=} \sum_{i=0}^I a_i. \quad (3.7)$$

We denote the resulting MAP predictors for $0 \leq j \leq I$ by

$$\hat{\mu}_i^{\text{MAP}*}(\mathbf{a}) = \hat{\mu}_i^{\text{MAP}*}(\mathbf{a}, \mathbf{m}, \varphi) \quad \text{and} \quad \hat{\gamma}_j^{\text{MAP}*}(\mathbf{a}) = \hat{\gamma}_j^{\text{MAP}*}(\mathbf{a}, \mathbf{m}, \varphi),$$

where $\hat{\gamma}_j^{\text{MAP}*}(\mathbf{a})$ is obtained from (3.3).

Remarks 3.5

- Predictors in the gamma priors Model 3.1 are denoted by a superscript “*”, e.g. $\hat{\gamma}_j^*$, whereas predictors in the uniform priors Model 2.1 are simply denoted by $\hat{\beta}_j$ and $\hat{\gamma}_j$, respectively, depending on the parametrization (2.5).
- Note that the MAP predictors can now easily be found by a simple (one-dimensional) root searching algorithm (it only depends on one single parameter $\mu > 0$, see (3.7)). This is slightly more involved than the closed form solution (2.10) in the uniform prior case, but it is a lot simpler than the multi-dimensional generalized linear model (GLM) claims reserving problems where one either uses the Newton-Raphson algorithm or Fisher’s scoring method to find the roots for the multidimensional problems, see for example Chapter 6 in Wüthrich & Merz (2008).
- For the special case of $m_i \equiv m$ and $a_i \equiv a$ we obtain a closed form solution. Equation (3.5) implies for constant m_i and a_i

$$\mu_i = \frac{C_{i,I-i}^* + a\varphi}{\sum_{k=0}^{i-1} (C_{k,I-i}^* + a\varphi)} \sum_{k=0}^{i-1} \mu_k.$$

The normalization condition (3.6) then provides

$$\begin{aligned} m(I+1) &\stackrel{!}{=} \sum_{i=0}^I \mu_i = \sum_{i=0}^I \frac{C_{i,I-i}^* + a\varphi}{\sum_{k=0}^{i-1} (C_{k,I-i}^* + a\varphi)} \sum_{k=0}^{i-1} \mu_k \\ &= \left(\frac{C_{I,0}^* + a\varphi}{\sum_{k=0}^{I-1} (C_{k,0}^* + a\varphi)} + 1 \right) \sum_{i=0}^{I-1} \mu_i = \dots = \prod_{j=0}^{I-1} \frac{\sum_{k=0}^{I-j} (C_{k,j}^* + a\varphi)}{\sum_{k=0}^{I-j-1} (C_{k,j}^* + a\varphi)} \mu_0. \end{aligned} \quad (3.8)$$

Hence, from this we can explicitly calculate the MAP predictor

$$\hat{\mu}_0^{MAP*}(\mathbf{a}) = m(I+1) \left(\prod_{j=0}^{I-1} \frac{\sum_{k=0}^{I-j} (C_{k,j}^* + a\varphi)}{\sum_{k=0}^{I-j-1} (C_{k,j}^* + a\varphi)} \right)^{-1},$$

and the iteration then provides the remaining MAP predictors.

- We can now study the MAP predictors as a function of the *degree of information* \mathbf{a} contained in the prior estimates m_i , in particular, we obtain a smoothed claims development pattern $\hat{\gamma}_j^{MAP*}(\mathbf{a})$, where the degree of smoothing depends on \mathbf{a} .

The MAP predictor for the outstanding loss liabilities of accident year $i > 0$ is then given by

$$\hat{R}_i^{MAP*}(\mathbf{a}) = \hat{\mu}_i^{MAP*}(\mathbf{a}) \sum_{j=I-i+1}^I \hat{\gamma}_j^{MAP*}(\mathbf{a}). \quad (3.9)$$

3.2 Non-informative prior case

For the non-informative prior case we let $a_i \rightarrow 0$ for all i . The posterior density (likelihood function) in Model 3.1 for non-informative priors is then asymptotically given by

$$u(\theta|\mathcal{D}_I) \propto \prod_{i+j \leq I} \exp\left\{-\frac{\mu_i \gamma_j}{\varphi}\right\} \left(\mu_i \gamma_j\right)^{\frac{X_{ij}}{\varphi}} \prod_{i=0}^I \mu_i^{-1} \prod_{j=0}^I \gamma_j^{-1}. \quad (3.10)$$

There are two important observations: (i) The non-informative prior case in Model 3.1 has exactly the same posterior density as the non-informative prior case in Model 2.1 “under the change of variables”, see (3.10) and (2.6) for $m, b \rightarrow \infty$. Therefore, the predictive posterior distributions of the outstanding loss liabilities R_i in these two non-informative priors models will coincide as well as their MAP predictors. (ii) Note that in the case (3.10) the last terms on the left-hand side and the right-hand side of (3.2) disappear. Therefore, we are left with

$$\mu_i \sum_{j=0}^{I-i} \gamma_j = \sum_{j=0}^{I-i} X_{ij}^*, \quad \text{for } i = 0, \dots, I, \quad (3.11)$$

$$\gamma_j \sum_{i=0}^{I-j} \mu_i = \sum_{i=0}^{I-j} X_{ij}^*, \quad \text{for } j = 0, \dots, I. \quad (3.12)$$

Similar to the solution of (2.1)–(2.2) we find the following solutions to (3.11)–(3.12)

$$\hat{\gamma}_j^{\text{MAP}*}(0) = e^K \left(1 - \frac{1}{\hat{f}_{j-1}^*}\right) \prod_{k=j}^{I-1} \frac{1}{\hat{f}_k^*} \quad \text{and} \quad \hat{\mu}_i^{\text{MAP}*}(0) = \frac{\sum_{j=0}^{I-i} X_{ij}^*}{\sum_{j=0}^{I-i} \hat{\gamma}_j^{\text{MAP}*}(0)},$$

for any positive constant e^K and CL factors \hat{f}_j^* for the transformed observations $C_{i,j+1}^*$

$$\hat{f}_j^* = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}^*}{\sum_{i=0}^{I-j-1} C_{i,j}^*}.$$

Therefore, we obtain a CL model for the incremental claims X_{ij}^* . However, in this case we can find a “natural” normalizing constant e^K . Theorem 3.4 implies

$$\mu_i(0) = \lim_{a \rightarrow 0} \mu_i(a) = \frac{C_{i,I-i}^*}{\sum_{k=0}^{i-1} C_{k,I-i}^*} \sum_{k=0}^{i-1} \mu_k(0). \quad (3.13)$$

Proposition 3.6 *In Model 3.1 equations (3.2)–(3.3) imply*

$$\lim_{a \rightarrow 0} \hat{\mu}_0^{\text{MAP}*}(a) = (I+1) \left[\frac{1}{m_0} + \sum_{j=0}^{I-1} \frac{1}{m_{I-j}} \frac{C_{I-j,j}^*}{\sum_{k=0}^{I-j-1} C_{k,j}^*} \prod_{n=j+1}^{I-1} \frac{\sum_{m=0}^{I-n} C_{m,n}^*}{\sum_{m=0}^{I-n-1} C_{m,n}^*} \right]^{-1}.$$

Therefore, Proposition 3.6 provides a natural scaling constant $e^K > 0$ if we let the degree of information a converge to 0.

3.3 Strong prior case

For the strong prior case we let $a_i \rightarrow \infty$ for all i and obtain from (3.2)–(3.3)

$$\begin{aligned} \frac{\mu_i}{m_i} &= 1, & \text{for } i = 0, \dots, I, \\ \gamma_j \sum_{i=0}^{I-j} \mu_i &= \sum_{i=0}^{I-j} X_{i,j}^*, & \text{for } j = 0, \dots, I. \end{aligned}$$

Therefore, $\hat{\mu}_i^{\text{MAP}*}(\infty) = m_i$ and

$$\hat{\gamma}_j^{\text{MAP}*}(\infty) = \sum_{i=0}^{I-j} X_{i,j}^* / \sum_{i=0}^{I-j} m_i.$$

In this case we can explicitly calculate the posterior distributions of γ_j , given \mathcal{D}_I . These posterior distributions are independent with

$$\gamma_j | \mathcal{D}_I \stackrel{(d)}{\sim} \Gamma \left(\sum_{i=0}^{I-j} X_{i,j} / \varphi, \sum_{i=0}^{I-j} m_i / \varphi \right). \quad (3.14)$$

This immediately implies that

$$\hat{\gamma}_j^{\text{MMSE}*}(\infty) = \frac{\sum_{i=0}^{I-j} X_{i,j}}{\sum_{i=0}^{I-j} m_i}, \quad (3.15)$$

and the bias of the MAP predictor of γ_j is given by

$$\hat{\gamma}_j^{\text{MMSE}*}(\infty) - \hat{\gamma}_j^{\text{MAP}*}(\infty) = \frac{\sum_{i=0}^{I-j} X_{i,j} - \sum_{i=0}^{I-j} X_{i,j}^*}{\sum_{i=0}^{I-j} m_i} = \frac{\varphi}{\sum_{i=0}^{I-j} m_i}.$$

Therefore, in the strong prior case we obtain closed form posterior distributions which allow for an analytical analysis of the model, both for the MAP predictor

$$\hat{R}_i^{\text{MAP}*}(\infty) = m_i \sum_{j=I-i+1}^I \hat{\gamma}_j^{\text{MAP}*}(\infty),$$

and the MMSE predictor

$$\hat{R}_i^{\text{MMSE}*}(\infty) = m_i \sum_{j=I-i+1}^I \hat{\gamma}_j^{\text{MMSE}*}(\infty).$$

3.4 Link to the Bornhuetter & Ferguson (1972) method

The BF method, as applied in practice, uses as claims development pattern γ_j the one implied by the CL factor estimates given in (2.10). Therefore, the classical BF predictor for the outstanding loss

liabilities is given by

$$\hat{R}_i^{BF} = m_i \frac{\sum_{j=i+1}^I e^{\hat{\beta}_j^{MAP}}}{\sum_{j=0}^I e^{\hat{\beta}_j^{MAP}}}, \quad (3.16)$$

where the $\hat{\beta}_j^{MAP}$'s solve (2.1)–(2.2). \hat{R}_i^{BF} exactly corresponds to the BF predictor studied in Alai *et al.* (2009).

Mack (2008) provides a different BF predictor where he uses a different method for the estimation of the claims development pattern γ_j . We include a comparison of the results with two versions of the Mack (2008) method. In the first case we define the raw pattern, see formula (3) in Mack (2008),

$$\hat{\gamma}_j^{raw} = \frac{\sum_{i=0}^{I-j} X_{i,j}}{\sum_{i=0}^{I-j} m_i} = \hat{\gamma}_j^{MMSE*}(\infty).$$

This pattern is not normalized, i.e. does not add up to 1. Therefore, we can also study a second development pattern defined by

$$\hat{\gamma}_j^{norm} = \frac{\hat{\gamma}_j^{raw}}{\sum_{j=0}^I \hat{\gamma}_j^{raw}}.$$

We then define similar to Mack (2008)

$$\hat{R}_i^{Mack1} = m_i \sum_{j=i+1}^I \hat{\gamma}_j^{raw} = \hat{R}_i^{MMSE*}(\infty) \quad \text{and} \quad \hat{R}_i^{Mack2} = m_i \sum_{j=i+1}^I \hat{\gamma}_j^{norm}.$$

These BF predictors \hat{R}_i^{BF} , \hat{R}_i^{Mack1} and \hat{R}_i^{Mack2} can now be compared to the CL predictor $\hat{R}_i^{CL} = \hat{R}_i^{MAP}$ as well as to the MAP predictors $\hat{R}_i^{MAP*}(\mathbf{a})$, for $a_i \in [0, \infty]$ and the corresponding MMSE predictors. In this spirit, the Bayesian predictors can be viewed as BF predictors where a_i determines the degree of information contained in the prior value m_i . These predictions and estimators are compared in Section 5.

Moreover, $\hat{\gamma}_j^{MAP*}(\mathbf{a})$ and $\hat{\gamma}_j^{MMSE*}(\mathbf{a})$ can be viewed as smoothed claims development patterns where we account for the prior information m_i according to its degree of information a_i for smoothing.

4 Bias, prediction uncertainty and MCMC

4.1 Gibbs sampler

In general, Models 2.1 and 3.1 do not allow for analytical calculations of the posterior distributions. In most cases the posterior distribution of the parameters can only be determined up to the normalizing constant. This is then the ideal situation to apply MCMC simulation methods which provide empirical posterior distributions. These empirical posterior distributions then allow for the calculation of claims reserves, cash flows and any desirable risk measure. For an introduction to MCMC methods we refer to Gilks *et al.* (1996), Asmussen & Glynn (2007) and Spiegelhalter *et al.* (1995, 2002). We mention that in recent actuarial literature MCMC methods

became rather popular, see e.g. Scollnik (2001) and the literature therein, England & Verrall (2002, 2006) and Section 4.4 in Wüthrich & Merz (2008).

Here, we use the Gibbs sampler, see Gilks *et al.* (1996), page 12. The Gibbs sampler is a simplified version of the single-component Metropolis-Hastings algorithm (Metropolis *et al.*, 1953, Hastings, 1970). Our aim is to sample from the posterior density $u(\theta|\mathcal{D}_I)$ with $\theta = (\mu, \gamma) = (\mu_0, \dots, \mu_I, \gamma_0, \dots, \gamma_I)$, see (3.1). This posterior density has the special property that

$$u(\mu|\mathcal{D}_I, \gamma) \text{ are independent gamma densities with parameters } a_i + \sum_{j=0}^{I-i} \frac{X_{ij}}{\varphi} \text{ and } \frac{a_i}{m_i} + \sum_{j=0}^{I-i} \frac{\gamma_j}{\varphi},$$

$$u(\gamma|\mathcal{D}_I, \mu) \text{ are independent gamma densities with parameters } b + \sum_{i=0}^{I-j} \frac{X_{ij}}{\varphi} \text{ and } \frac{b}{c_j} + \sum_{i=0}^{I-j} \frac{\mu_i}{j}.$$

Thus, from these conditional posterior densities $u(\mu|\mathcal{D}_I, \gamma)$ and $u(\gamma|\mathcal{D}_I, \mu)$ we can directly sample from. The Gibbs sampler then goes as follows:

1. Initialize $\Theta^{(0)} = (\mu^{(0)}, \gamma^{(0)})$.
2. For $t \geq 1$ do
 - (a) generate $\mu^{(t)} \sim u(\cdot|\mathcal{D}_I, \gamma^{(t-1)})$;
 - (b) generate $\gamma^{(t)} \sim u(\cdot|\mathcal{D}_I, \mu^{(t)})$;
 - (c) set $\Theta^{(t)} = (\mu^{(t)}, \gamma^{(t)})$.

Then, this algorithm provides a Markov chain $(\Theta^{(t)})_{t \geq 0}$ whose stationary limit distribution is given by $u(\theta|\mathcal{D}_I)$, see Gilks *et al.* (1996) and Asmussen & Glynn (2007).

4.2 Empirical distribution from Gibbs sampling

Using the Gibbs sampler we obtain (after burn-in T) an empirical distribution from the sample

$$(\Theta^{(t)} = (\mu^{(t)}, \gamma^{(t)}))_{t > T} = \left((\mu_0^{(t)}, \dots, \mu_I^{(t)}, \gamma_0^{(t)}, \dots, \gamma_I^{(t)}) \right)_{t > T}$$

which is an estimator for the posterior distribution $u(\cdot|\mathcal{D}_I)$. Therefore, we estimate the MMSE predictor \hat{R}_i^{MMSE} by the sample mean

$$\hat{\hat{R}}_i^{MMSE} = \frac{1}{\hat{T} - T} \sum_{t=T+1}^{\hat{T}} \sum_{j=I-i+1}^I \mu_i^{(t)} \gamma_j^{(t)}.$$

To indicate that this is the sample mean we use two hats in the notation. The conditional MSE of the MMSE predictor is estimated similarly. Note that

$$\begin{aligned} \text{mse}_{R_i|\mathcal{D}_I}(\hat{\hat{R}}_i^{MMSE}) &= \text{Var}(R_i|\mathcal{D}_I) \\ &= E[\text{Var}(R_i|\mathcal{D}_I, \Theta)|\mathcal{D}_I] + \text{Var}(E[R_i|\mathcal{D}_I, \Theta]|\mathcal{D}_I) \\ &= \varphi E[R_i|\mathcal{D}_I] + \text{Var}\left(\sum_{j=I-i+1}^I \mu_i \gamma_j \middle| \mathcal{D}_I\right). \end{aligned} \quad (4.1)$$

Therefore, we get the estimator

$$\widehat{\text{mse}}_{R_i|\mathcal{D}_I}(\widehat{R}_i^{\text{MMSE}}) = \widehat{\varphi}_{\widehat{R}_i}^{\text{MMSE}} + \frac{1}{\widetilde{T} - T} \sum_{t=T+1}^{\widetilde{T}} \left(\sum_{j=L-i+1}^I \mu_i^{(t)} \gamma_j^{(t)} \right)^2 - \left(\widehat{R}_i^{\text{MMSE}} \right)^2.$$

Remarks 4.1

- We would like to emphasize that using the Gibbs sampler we do not only estimate the conditional MSE. The Gibbs sampler provides an approximation to the full posterior distribution $u(\cdot|\mathcal{D}_I)$ and one can calculate any desirable risk measure.
- The empirical sample $(\Theta^{(t)})_{t>T}$ allows for the simulation of the payments $X_{i,j}$; for any $t > T$ we may sample for $i + j > I$

$$X_{i,j}^{(t)} / \varphi \stackrel{(d)}{\sim} \text{Poi}(\mu_i^{(t)} \gamma_j^{(t)} / \varphi). \quad (4.2)$$

This provides the simulated cash flows. The sampled outstanding loss liabilities R_i are then obtained by

$$R_i^{(t)} = \sum_{j=L-i+1}^I X_{i,j}^{(t)}. \quad (4.3)$$

The sample $(R_i^{(t)})_{t>T}$ then provides the empirical posterior distribution of R_i given \mathcal{D}_I , see also Figure 2. Moreover, it also allows for the direct estimation of (4.1), simply by calculating the sample variance of the simulated values.

5 Example

5.1 Univariate example

Before we start with a real data example (in the next subsection) we illustrate the behaviour of the MAP and the MMSE predictors in a univariate example. This example highlights the importance of the choice of the prior distribution, the link function and its implications.

Assume conditionally, given Λ , X_1, \dots, X_n, X_{n+1} are i.i.d. Poisson distributed with parameter Λ . We assume that X_1, \dots, X_n are observed and we would like to make Bayesian inference for Λ and X_{n+1} .

We now make different choices for the distribution Λ :

Case 1. $\Gamma = \log(\Lambda)$ has a non-informative uniform prior distribution. In that case the posterior distribution of Γ , given $\mathbf{X} = (X_1, \dots, X_n)$, is given by

$$u(\Gamma|\mathbf{X}) \propto e^{-ne^\Gamma} (e^\Gamma)^{\sum_{i=1}^n X_i}. \quad (5.1)$$

This implies

$$\widehat{e^\Gamma}^{\text{MAP}} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad \widehat{\Lambda}^{\text{MMSE}} = E[e^\Gamma|\mathbf{X}] = \frac{1}{n} \sum_{i=1}^n X_i.$$

In this case the MAP and the MMSE predictors coincide.

Case 2. We make the same assumptions as in Case 1 but we do a change of variable in (5.1). We set $\Lambda = e^\Gamma$ this provides posterior density

$$u(\Lambda|\mathbf{X}) \propto e^{-n\Lambda} \Lambda^{\sum_{i=1}^n X_i - 1}.$$

This implies

$$\hat{\Lambda}^{MAP} = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \quad \text{and} \quad \hat{\Lambda}^{MMSE} = \frac{1}{n} \sum_{i=1}^n X_i.$$

That is, we obtain $\hat{\Lambda}^{MAP} < \hat{\Lambda}^{MMSE}$. This shows that the MAP predictors are not invariant under re-parametrization and therefore are often not appropriate. This is well-known in Bayesian theory, see for example Smith (1998).

Case 3. Λ has a non-informative gamma prior distribution. In that case the posterior distribution of Λ , given \mathbf{X} , has exactly the same form as in Case 2 and therefore we obtain the same inference picture as in Case 2.

Case 4. Λ has the non-informative Jeffrey's prior distribution $\lambda^{-1/2}$. In that case the posterior distribution of Λ , given $\mathbf{X} = (X_1, \dots, X_n)$, is given by

$$u(\Lambda|\mathbf{X}) \propto e^{-n\Lambda} \Lambda^{\sum_{i=1}^n X_i - 1/2}.$$

Jeffrey's non-informative priors are often used because they have invariance properties under parameter transformations. This implies

$$\hat{\Lambda}^{MAP} = \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2n} \quad \text{and} \quad \hat{\Lambda}^{MMSE} = \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{2n}.$$

In this paper we do not further investigate Jeffrey's priors.

Conclusion. The MMSE predictor $\hat{\Lambda}^{MMSE}$ has always minimal posterior variance and is invariant under re-parametrization. Therefore the optimal Bayesian predictor for X_{n+1} , given \mathcal{D}_I , is always given by

$$E[X_{n+1}|\mathbf{X}] = E[E[X_{n+1}|\Lambda, \mathbf{X}]|\mathbf{X}] = E[\Lambda|\mathbf{X}] = \hat{\Lambda}^{MMSE}.$$

5.2 Real data example

We revisit the BF example given in Tables 2.2–2.4 of Wüthrich & Merz (2008) (this is the example also considered in the BF analysis in Alai *et al.*, 2009)), see Table 1. We analyze this data set for non-informative uniform priors according to Model 2.1 and for gamma priors according to Model 3.1. In order to compare the results to the results in Wüthrich & Merz (2008) and Alai *et al.* (2009) we choose a fixed plug-in estimate $\varphi = 14,714$.

5.2.1 Non-informative priors and the CL method

In this subsection we study Model 2.1 with non-informative uniform priors and log link as well as Model 3.1 with non-informative gamma priors. The Gibbs sampler allows us to numerically calculate the MMSE predictors

$$\hat{R}^{MMSE} = \sum_i \hat{R}_i^{MMSE} \quad \text{and} \quad \hat{R}^{MMSE*} = \sum_i \hat{R}_i^{MMSE*},$$

Table 1. Observed incremental claims $X_{i,j}$, $i + j \leq I$, and prior values m_i .

i/j	0	1	2	3	4	5	6	7	8	9	m_i
0	5,946,975	3,721,237	895,717	207,760	206,704	62,124	65,813	14,850	11,130	15,813	11,653,101
1	6,346,756	3,246,406	723,222	151,797	67,824	36,603	52,752	11,186	11,646		11,367,306
2	6,269,090	2,976,233	847,053	262,768	152,703	65,444	53,545	8,924			10,962,965
3	5,863,015	2,683,224	722,532	190,653	132,976	88,340	43,329				10,616,762
4	5,778,885	2,745,229	653,894	273,395	230,288	105,224					11,044,881
5	6,184,793	2,828,338	572,765	244,899	104,957						11,480,700
6	5,600,184	2,893,207	563,114	225,517							11,413,572
7	5,288,066	2,440,103	528,043								11,126,527
8	5,290,793	2,357,936									10,986,548
9	5,675,568										11,618,437

Table 2. Claims reserves predictors with corresponding conditional $\text{MSEP}^{1/2}$ in Model 2.1 (with non-informative uniform priors and log link) and in Model 3.1 (with non-informative gamma priors). The figures in brackets are obtained by Gibbs sampling, the others are exact. The results are compared to the frequentist's CL model of England & Verrall (2002) and of Mack (1993). Note that Mack (1993) is a rather different model, so we include Mack's (1993) results only for comparison purposes.

	claims reserves	posterior bias term	$\text{MSEP}^{1/2}$
Model 2.1 (non-informative uniform priors with log link)			
$\hat{R}^{\text{MAP}} = \hat{R}^{\text{CL}}$	6,047,059	(-2,339)	(430,166)
\hat{R}^{MMSE}	(6,049,398)		(430,160)
Model 3.1 (non-informative gamma priors)			
$\hat{R}^{\text{MAP}*}$	5,783,089	(-266,229)	(505,881)
$\hat{R}^{\text{MMSE}*}$	(6,049,398)		(430,160)
frequentist's CL model of England & Verrall (2002) and of Mack (1993)			
\hat{R}^{CL} from ODP, England & Verrall (2002)	6,047,059		429,891
\hat{R}^{CL} from Mack (1993)	6,047,059		462,960

for Model 2.1 (with $m, b \rightarrow \infty$) and Model 3.1 (with $a_i \equiv a \rightarrow 0$ and $b \rightarrow 0$). In Model 2.1 the posterior density is then given by (2.5)–(2.6) with $m, b \rightarrow \infty$. In Model 3.1 the posterior density is then given by (3.10).

Note that for these two non-informative prior cases the posterior densities coincide, see (2.6) and (3.10). Therefore, we only need to run one Gibbs simulation to solve both of these two cases numerically.

We have used the Gibbs sampler and we have run 1,000,000 simulations after the subtraction of burn-in costs $T = 100,000$. This provided the empirical posterior distribution of the parameters $(\mu_0^{(t)}, \dots, \mu_I^{(t)}, \gamma_0^{(t)}, \dots, \gamma_I^{(t)})_{t=100,001, \dots, 1,100,000}$ from which the MMSE predictors and their empirical uncertainty $\widehat{\text{mse}}_{R|D_I}(\cdot)$ were provided, see Section 4.2. For the estimation of the prediction uncertainty of the MAP predictor we have used formula (2.8). The results are presented in Table 2.

Observations 5.1

- We observe that the predictors of the outstanding loss liabilities are all rather similar in these non-informative prior situations. The MAP predictor $\hat{R}^{MAP} = 6,047,059$ coincides with the CL reserves \hat{R}^{CL} and it is also in line with the MMSE predictors $\hat{R}^{MMSE} = \hat{R}^{MMSE*} = 6,049,398$ obtained by Gibbs sampling. Only the MAP predictor in the non-informative gamma priors Model 3.1 gives a prediction $\hat{R}^{MAP*} = 5,783,089$ that deviates from the others. This prediction seems too low and moreover, as mentioned in Remarks 2.3, the MAP predictor is not invariant under re-parametrization. Therefore its use is questionable.
- Note that although the MAP predictors for Models 2.1 and 3.1 (with non-informative priors) are different, the distributions of the reserves are identical since they are from the same Gibbs simulation. This highlights the danger of focusing solely on the MAP predictors, and not on the distribution.
- Prediction uncertainties in terms of the conditional MSEP: We compare our Bayesian calculations to the frequentist's estimates found in the literature: (i) ODP (constant scale) analytical approximation using asymptotic normality of MLEs, see Section 7.2 in England & Verrall (2002), (ii) distribution-free CL method, see Mack (1993):

$$\widehat{\text{mse}}_{R|D_i}(\hat{R}^{CL}) = \begin{cases} \text{ODP (constant scale) approximation, England \& Verrall (2002),} \\ \text{according to Mack's distribution – free CL model (1993).} \end{cases}$$

We observe that our Bayesian models provide a prediction uncertainty in the range of 430,000. This is very similar to the estimate of England & Verrall (2002) in the asymptotic normality approximation. Mack (1993)'s model is a rather different model, therefore we include Mack's (1993) results only for comparison purposes.

- The Bayesian models now have the advantage that they provide the full posterior parameter distributions. Therefore, we can calculate the predictive distribution of the outstanding loss liabilities (not only the claims reserves and the conditional MSEP). This is further outlined below.

5.2.2 Informative gamma priors

We turn to Model 3.1 (gamma priors) with informative priors, that is, we implement prior knowledge about the exposure parameters μ_i . We choose the degree of information a constant for all accident years, i.e. $a_i \equiv a \in [0, \infty]$. Then the MAP predictors in Model 3.1 are given by

$$\hat{R}^{MAP*}(a) = \sum_i \hat{R}_i^{MAP*}(a) = \sum_{i+j>I} \hat{\mu}_i^{MAP*}(a) \hat{\gamma}_j^{MAP*}(a).$$

These are calculated by the root searching algorithm given in (3.7) for $a \in (0, \infty)$, the cases $a = 0$ and $a = \infty$ can be solved explicitly. Figure 1 gives the MAP predictors $\hat{R}^{MAP*}(a)$ for different degrees of information $a \in [0, \infty]$. We see that in our case the claims reserves $\hat{R}^{MAP*}(a)$ are an increasing function in the degree of information a . This comes from the fact that the prior estimates m_i are rather conservative (this will be further highlighted below).

Next, we determine the MMSE predictors and the prediction uncertainties in the gamma priors model. Therefore we again apply the Gibbs sampler. After subtracting the burn-in costs $T = 100,000$ we again simulate 1,000,000 samples. The results are provided in Table 3.

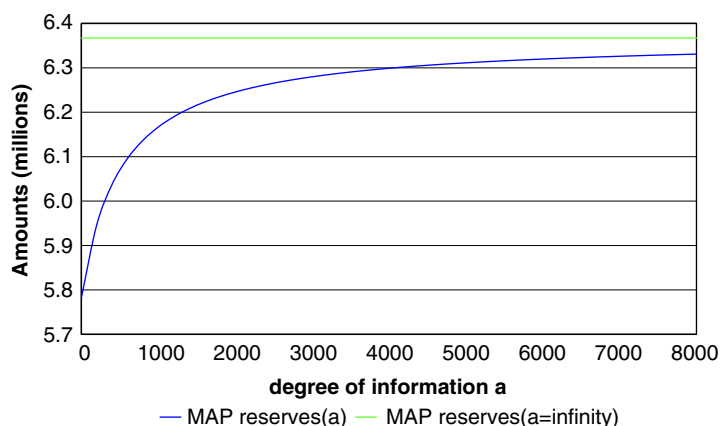


Figure 1. MAP predictors $\hat{R}^{MAP*}(a)$ for different degrees of information $a \in [0, \infty]$.

Table 3. Claims reserves predictors with corresponding conditional $MSEP^{1/2}$ in the gamma Model 3.1 for different degrees of information a . The figures in brackets are obtained by Gibbs sampling.

	claim reserves	posterior bias term	$MSEP^{1/2}$
$\hat{R}^{MAP*}(a = 0)$	5,783,089	(-266,229)	(505,881)
$\hat{R}^{MMSE*}(a = 0)$	(6,049,398)		(430,160)
$\hat{R}^{MAP*}(a = 100)$	5,878,911	(-266,615)	(499,621)
$\hat{R}^{MMSE*}(a = 100)$	(6,145,526)		(422,526)
$\hat{R}^{MAP*}(a = \infty)$	6,367,134	-276,913	482,406
$\hat{R}^{MMSE*}(a = \infty)$	6,644,047		395,012

Observations 5.2

- The first observation is that in the Gibbs sampler we obtain long-range dependencies for later development periods. This comes from the fact that we have large variances (for non-informative priors) and only a few observations. Therefore, we need many simulations (\hat{T} large) for the convergence of the empirical mean.
- Similar to the non-informative gamma prior case we see substantial posterior bias terms in the MAP predictors. This comes from the fact that the dispersion φ is fairly large compared to the incremental payments $X_{i,j}$ in later development periods. For $a = \infty$, for example, this results in $\hat{\gamma}_9^{MAP*}(\infty) = 0.01\%$ and $\hat{\gamma}_9^{MMSE*}(\infty) = 0.14\%$ which explains the posterior bias terms. This again indicates that the MAP predictors should not be used.
- We see that the MMSE predictors are increasing in the degree of information a . This comes from the fact that the prior means m_i were chosen rather conservative and the more weight we give to these conservative prior means the more the MMSE predictors increase.
- The conditional MSEPs are decreasing in the degree of information a . This is rather obvious because the less uncertainty we have in the prior distributions the less prediction uncertainty we obtain. We see that the conditional $MSEP^{1/2}$ decreases from 430,160 to 395,012.

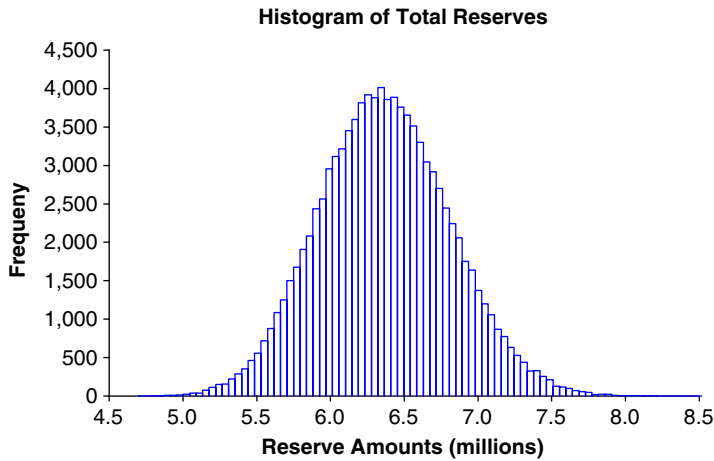


Figure 2. Histogram for $R|\mathcal{D}_I$ in Model 3.1 for $a = 100$ from 100,000 simulations.

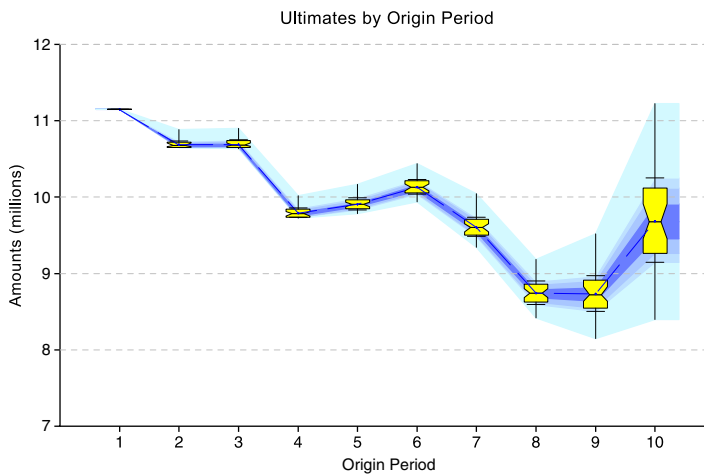


Figure 3. Confidence intervals of $R_i|\mathcal{D}_I$, $i = 1, \dots, 10$, in Model 3.1 for $a = 100$. The different quantiles correspond to: minimum, 5th percentile, 10th percentile, 25th percentile, 50th percentile, 75th percentile, 90th percentile, 95th percentile and maximum.

As mentioned above, we obtain the full posterior distribution from the Gibbs sampler for the outstanding loss liabilities $R = \sum_i R_i$, conditional on \mathcal{D}_I , see (4.3). We consider in Model 3.1 the case $a = 100$. The histogram of the total reserves from 100,000 simulation of the outstanding loss liabilities R is given in Figure 2.

This empirical distribution now allows for the estimation of any risk measure, not only the conditional MSE. Moreover, we can also plot confidence intervals, for example, in Figure 3 we show the confidence intervals per accident year i . As expected, we observe that the uncertainty in

old accident years is rather low, because they are already well developed, whereas for younger accident years we obtain bigger ranges.

The Gibbs sampler not only provides the conditional distribution of the outstanding loss liabilities R_i , given \mathcal{D}_I , but we also obtain the conditional distribution of the cash flows $X_{i,j}$, conditionally given \mathcal{D}_I . From these cash flows we can determine how the uncertainty evolves over time (over the development years). In Figure 4 we show the development of the uncertainty over time for the youngest accident year. We see that the payment for the first development year is given (contained in \mathcal{D}_I). This is why there is no uncertainty at time 1. After this first development year we obtain the corresponding confidence intervals. Figure 5 describes the same development uncertainty but for the

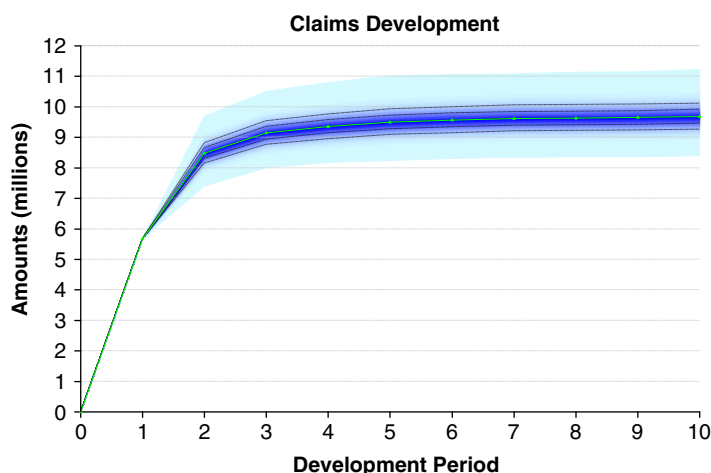


Figure 4. Development of $C_{i,j}$, $i = I$ and $j \geq 0$, over time in Model 3.1 for $a = 100$. The different gray scales correspond to: 10th percentile, 25th percentile, 50th percentile, 75th percentile, 90th percentile.

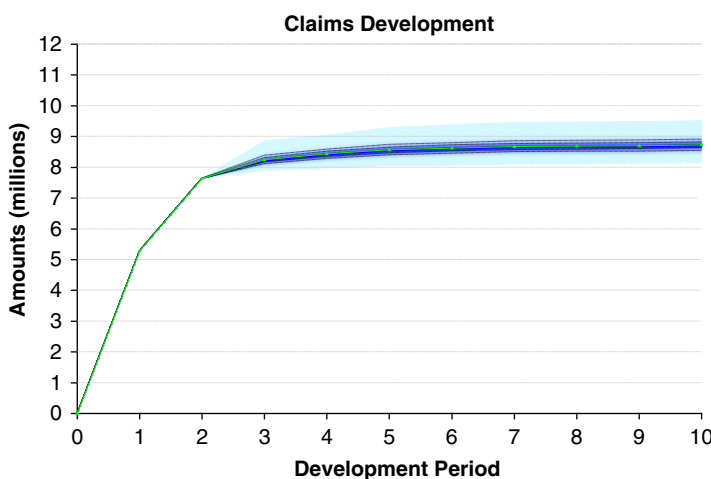


Figure 5. Development of $C_{i,j}$, $i = I-1$ and $j \geq 0$, over time in Model 3.1 for $a = 100$. The different gray scales correspond to: 10th percentile, 25th percentile, 50th percentile, 75th percentile, 90th percentile.

second youngest accident year. Of course, we observe a smaller uncertainty, because we have more observations compared to the case in Figure 4.

5.2.3 Strong gamma prior case and the BF method

Finally, we compare the gamma priors Model 3.1 with strong priors for μ_i to the classical BF predictor. In the literature the classical BF predictor is given by (3.16). We also compare the classical BF predictor to the BF predictors obtained from Mack (2008):

$$\begin{aligned}\hat{R}^{Mack1} &= \sum_i \hat{R}_i^{Mack1} = \sum_{i+j-1} m_i \hat{\gamma}_j^{raw} \quad (\text{BF-Mack predict or from raw pattern}), \\ \hat{R}^{Mack2} &= \sum_i \hat{R}_i^{Mack2} = \sum_{i+j-1} m_i \hat{\gamma}_j^{norm} \quad (\text{BF-Mack predict or from normalized pattern}).\end{aligned}$$

For the calculation of the prediction uncertainty of the BF predictor there are different methods in the literature: The conditional MSE $\widehat{\text{mse}}_{R|D_1}(\hat{R}^{BF})$ of the classical BF predictor is calculated with Alai *et al.* (2009) and the conditional MSEs $\widehat{\text{mse}}_{R|D_1}(\hat{R}^{Mack1})$ and $\widehat{\text{mse}}_{R|D_1}(\hat{R}^{Mack2})$ of the BF-Mack predictors are calculated according to Mack (2008). The results are presented in Table 4.

Observations 5.3

- The different BF predictors are rather diverse. This comes from the fact that the prior values m_i are too high, which has the rather unpleasant effect that we do not obtain reliable estimates for the claims development pattern γ_j . For the raw pattern we obtain $\sum_i \hat{\gamma}_j^{raw} = \sum_j \hat{\gamma}_j^{MAP*}(\infty) < 1$. If we normalize this raw pattern, we get predictors \hat{R}^{Mack2} and \hat{R}^{BF} that are too high. Also the non-normalized ones $\hat{R}^{MMSE*}(a = \infty)$ and \hat{R}^{Mack1} seem to be too high because of the large values of m_i .
- The predictors $\hat{R}^{MMSE*}(a = \infty)$ and \hat{R}^{Mack1} coincide because they use the same parameter estimates. However, the underlying reasoning is slightly different which can be seen in the prediction uncertainty. For the MMSE predictor $\hat{R}^{MMSE*}(a = \infty)$ there is no uncertainty in m_i (because we assume perfect information $a = \infty$), whereas in \hat{R}^{Mack1} we also add uncertainty to m_i .
- The gamma priors Model 3.1 is consistent in the sense that it also uses the prior knowledge on μ_i to estimate the claims development pattern γ_j (whereas the other BF methods are not). In this spirit our Bayes model should be preferred. Moreover, we also have the flexibility to attach credibility weights in terms of a to this prior knowledge which then results in Table 3.

Table 4. Claims reserves predictors with corresponding conditional MSE^{1/2} according to Model 3.1 with strong priors, Alai *et al.* (2009) and Mack (2008).

	claims reserves	MSE ^{1/2}
$\hat{R}^{MMSE*}(a = \infty)$ from Model 3.1	6,644,047	395,012
\hat{R}^{BF} from BF (1972) with Alai <i>et al.</i> (2009)	7,356,578	471,973
\hat{R}^{Mack1} from Mack (2008) raw pattern	6,644,047	539,678
\hat{R}^{Mack2} from Mack (2008) normalized pattern	7,505,455	726,531

6 Conclusions

The Bayesian ODP claims reserving model with uniform priors and log link (Model 2.1) and with gamma priors (Model 3.1) give mathematically consistent ways to estimate claims reserves in the Bornhuetter & Ferguson (1972) spirit:

- they use prior knowledge m_i for the expected ultimate claim;
- they combine the prior knowledge m_i with an estimated claims development pattern $\hat{\gamma}_j$ to obtain the reserves;
- this claims development pattern is estimated using a credibility weighted average between the observations \mathcal{D}_I and the prior knowledge m_i according to the degree of information \mathbf{a} contained in the prior knowledge. Complete prior knowledge ($\mathbf{a} = \infty$) leads to a BF model similar to Mack (2008), no prior knowledge ($\mathbf{a} = 0$) leads to the CL case, and for $\mathbf{a} \in (0, \infty)$ we can model any intermediate case.

The advantage of such full Bayesian models is that they allow for a complete analysis and for the calculation of any risk measure, whereas the frequentist's approaches (Alai *et al.* (2009) and Mack (2008)) need additional approximations for the determination of the conditional MSEF, and are unable to provide additional information such as predictive distributions of cash flows.

Limitations and outlook for further research. This paper only considers the ODP model with constant scale factor φ , and the BF model in the context of the CL model without a tail factor. In many cases the choice of a constant scale parameter φ should be checked. Often data suggests that φ_j depends on the development period j . Furthermore, it should be checked whether the conditional independence assumption between the $X_{i,j}$'s is appropriate and whether one should include tail factors beyond the latest development period.

References

- Alai, D.H., Merz, M. & Wüthrich, M.V. (2009). Mean square error of prediction in the Bornhuetter-Ferguson claims reserving method. *Annals of Actuarial Science*, **4/1**, 7–31.
- Asmussen, S. & Glynn, P.W. (2007). *Stochastic Simulation*. Springer.
- Bornhuetter, R.L. & Ferguson, R.E. (1972). The actuary and IBNR. *Proceedings CAS*, **LIX**, 181–195.
- England, P.D. & Verrall, R.J. (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal*, **8/3**, 443–518.
- England, P.D. & Verrall, R.J. (2006). Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science*, **1/2**, 221–270.
- Gilks, W.R., Richardson, S. & Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.
- Mack, T. (1991). A simple parametric model for rating automobile insurance or estimating IBNR claims reserves. *ASTIN Bulletin*, **21/1**, 93–109.
- Mack, T. (1993). Distribution-free calculation of the standard error of chain ladder reserve estimates. *ASTIN Bulletin*, **23/2**, 213–225.
- Mack, T. (2008). The prediction error of Bornhuetter/Ferguson. *ASTIN Bulletin*, **38/1**, 87–103.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal Chemical Physics*, **21/6**, 1087–1092.

- Saluz, A., Gisler, A. & Wüthrich, M.V (2011). Development pattern and prediction error for the stochastic Bornhuetter-Ferguson claims reserving model. *ASTIN Bulletin*, **41/2**, 279–313.
- Scollnik, D.P.M. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial Journal*, **5/2**, 96–125.
- Smith, R.L. (1998). Bayesian and frequentist approach to parametric predictive inference. In: *Bayesian statistics*, **6**, Bernardo, J.M., Berger, J.O., Dawid, A.P. & Smith, A.F.M. (eds.) Oxford University Press, 589–612.
- Spiegelhalter, D.J., Thomas, A., Best, N.G. & Gilks, W.R. (1995). *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.5*. MRC Biostatistics Unit, Cambridge.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal Royal Statistical Society*, **64**, 583–639.
- Verrall, R.J. (2004). A Bayesian generalized linear model for the Bornhuetter-Ferguson method of claims reserving. *North American Actuarial Journal*, **8/3**, 67–89.
- Wüthrich, M.V. & Merz, M. (2008). *Stochastic Claims Reserving Methods in Insurance*. Wiley.

A Proofs

Proof of Lemma 3.2. This is an immediate consequence of the assumptions and (3.2)–(3.3). \square

Lemma A.1 In Model 3.1 equations (3.2)–(3.3) imply for $j = 0, \dots, I$

$$\sum_{k=0}^{I-j} \mu_k \sum_{m=0}^j \gamma_m = \sum_{k=0}^{I-j} C_{k,j}^* + \varphi \sum_{k=0}^{I-j} a_k \left(1 - \frac{\mu_k}{m_k}\right).$$

Proof of Lemma A.1. We first prove the two statements (an empty sum is set equal to 0)

$$\sum_{k=0}^{I-j} C_{k,j}^* = \sum_{k=0}^{I-j} \mu_k \sum_{m=0}^j \gamma_m + \varphi \sum_{k=I-j+1}^I a_k \left(1 - \frac{\mu_k}{m_k}\right), \quad (\text{A.1})$$

$$\sum_{i=0}^I \frac{a_i \mu_i}{m_i} = \sum_{i=0}^I a_i. \quad (\text{A.2})$$

If we sum (3.2) over $i = 0, \dots, I$ and (3.3) over $j = 0, \dots, I$ we obtain

$$\sum_{i+j \leq I} \gamma_j \mu_i + \varphi \sum_{i=0}^I \frac{a_i \mu_i}{m_i} - \varphi \sum_{i=0}^I a_i = \sum_{i+j \leq I} X_{i,j}^* = \sum_{i+j \leq I} \gamma_j \mu_i.$$

This immediately implies statement (A.2). We now turn to (A.1). The proof is similar to the proof of Lemma 2.17 in Wüthrich & Merz (2008) and goes by induction.

We start with $j = 0$: using (3.3) in the second step we have

$$\sum_{k=0}^I C_{k,0}^* = \sum_{k=0}^I X_{k,0}^* = \gamma_0 \sum_{k=0}^I \mu_k.$$

This proves the claim for $j = 0$.

Induction step $j \rightarrow j+1$. We assume that the claim holds true for $j \leq I-1$, then we prove the claim for $j+1$:

$$\sum_{k=0}^{I-(j+1)} C_{k,j+1}^* = \sum_{k=0}^{I-j-1} \sum_{m=0}^{j+1} X_{k,m}^* = \sum_{k=0}^{I-j} \sum_{m=0}^j X_{k,m}^* - \sum_{m=0}^j X_{I-j,m}^* + \sum_{k=0}^{I-j-1} X_{k,j+1}^*.$$

To the first term on the right-hand side we apply the induction assumption

$$\sum_{k=0}^{I-j} \sum_{m=0}^j X_{k,m}^* = \sum_{k=0}^{I-j} C_{k,j}^* = \sum_{k=0}^{I-j} \mu_k \sum_{m=0}^j \gamma_m + \varphi \sum_{k=I-j+1}^I a_k \left(1 - \frac{\mu_k}{m_k}\right),$$

and to the second and third term (3.2) and (3.3), respectively. This implies

$$\begin{aligned} \sum_{k=0}^{I-(j+1)} C_{k,j+1}^* &= \sum_{k=0}^{I-j} \mu_k \sum_{m=0}^j \gamma_m + \sum_{k=I-j+1}^I \varphi a_k \left(1 - \frac{\mu_k}{m_k}\right) \\ &\quad - \left[\mu_{I-j} \sum_{k=0}^j \gamma_k + \varphi a_{I-j} \left(\frac{\mu_{I-j}}{m_{I-j}} - 1\right) \right] + \gamma_{j+1} \sum_{k=0}^{I-j-1} \mu_k \\ &= \sum_{k=0}^{I-(j+1)} \mu_k \sum_{m=0}^{j+1} \gamma_m + \varphi \sum_{k=I-(j+1)+1}^I a_k \left(1 - \frac{\mu_k}{m_k}\right). \end{aligned}$$

This proves (A.1). If we now combine (A.1) and (A.2) we obtain

$$\sum_{k=0}^{I-j} \mu_k \sum_{m=0}^j \gamma_m = \sum_{k=0}^{I-j} C_{k,j}^* - \varphi \sum_{k=I-j+1}^I a_k \left(1 - \frac{\mu_k}{m_k}\right) = \sum_{k=0}^{I-j} C_{k,j}^* + \varphi \sum_{k=0}^{I-j} a_k \left(1 - \frac{\mu_k}{m_k}\right).$$

This proves the claim. \square

Proof of Lemma 3.3. Choose $j \leq I-1$ then we have from Lemma A.1 and equation (3.2)

$$\begin{aligned} \sum_{k=0}^{I-j-1} \mu_k \sum_{m=0}^j \gamma_m &= \sum_{k=0}^{I-j} \mu_k \sum_{m=0}^j \gamma_m - \mu_{I-j} \sum_{m=0}^j \gamma_m \\ &= \sum_{k=0}^{I-j} C_{k,j}^* + \varphi \sum_{k=0}^{I-j} a_k \left(1 - \frac{\mu_k}{m_k}\right) - \sum_{k=0}^j X_{I-j,k}^* - a_{I-j} \varphi \left(1 - \frac{\mu_{I-j}}{m_{I-j}}\right) \quad (\text{A.3}) \\ &= \sum_{k=0}^{I-j-1} C_{k,j}^* + \varphi \sum_{k=0}^{I-j-1} a_k \left(1 - \frac{\mu_k}{m_k}\right). \end{aligned}$$

If we divide the equality in Lemma A.1 by (A.3) we obtain the claim. \square

Proof of Theorem 3.4. We solve (3.4) for μ_i . In a first step we obtain for $i = 1, \dots, I$

$$\mu_i \left[\sum_{k=0}^{i-1} C_{k,I-i}^* + \varphi \sum_{k=0}^{i-1} a_k \left(1 - \frac{\mu_k}{m_k}\right) \right] = \left(\sum_{k=0}^{i-1} \mu_k \right) \left[C_{i,I-i}^* + \varphi a_i \left(1 - \frac{\mu_i}{m_i}\right) \right].$$

Moreover, we have

$$\mu_i \left[\sum_{k=0}^{i-1} C_{k,l-i}^* + \varphi \sum_{k=0}^{i-1} a_k \left(1 - \frac{\mu_k}{m_k} \right) + \frac{a_i \varphi}{m_i} \sum_{k=0}^{i-1} \mu_k \right] = \left(\sum_{k=0}^{i-1} \mu_k \right) [C_{i,l-i}^* + a_i \varphi].$$

Therefore, if we divide by the bracket on the left-hand side we obtain

$$\mu_i = \frac{\left(\sum_{k=0}^{i-1} \mu_k \right) [C_{i,l-i}^* + a_i \varphi]}{\sum_{k=0}^{i-1} C_{k,l-i}^* + \varphi \sum_{k=0}^{i-1} \mu_k \left(\frac{a_k}{\mu_k} - \frac{a_k}{m_k} + \frac{a_i}{m_i} \right)}.$$

But then the first claim easily follows. The second claim was already proved in (A.2). \square

Proof of Proposition 3.6. The proof follows from the normalization condition (3.6) and (3.13) similar to the derivation (3.8). \square